公開

研 究 終 了 報 告 書

「学習問題の統合的帰着」

研究期間： 2020 年 11 月～2023 年 3 月
研 究 者： 末廣　大貴

1． 研究のねらい

学習問題は，しばしば分類問題，回帰問題，ランキング問題といったタスクの種類や，「弱教師あり」「教師なし」といったデータの条件などに基づいて体系化されている．それに伴い，多くの研究では**タスクのドメインごとに学習問題の定式化や解法アルゴリズムの解析が行われている**．しかし，実験的な解析のみが行われ，**理論解析が充分になされていないことも多い**．昨今の人工知能ブームに伴い，学習問題が増加の一途を辿る中，定式化は煩雑化しており，一般的な数理解析が行いづらいというのが最大の要因である．

理論解析を大幅に加速するためには，**これまでのタスク・ドメイン依存の個別解析から脱却し，**なんらかの方法で**統合的に解析を行っていく**が必要がある．ここで，ある複数の問題が「統合的に理論解析可能」というのは，それらがある1つの学習問題に帰着できるということを意味する．「学習問題の帰着」とは，ある学習問題を，別の学習問題に変換することで解法を導き出すことである．例えば，学習問題Ａが学習問題Ｂに帰着可能であれば，学習問題Ｂの解法アルゴリズムや，理論性能（計算量や汎化誤差など）を転用することができる．つまり，たとえ異なるタスク・データ条件であっても，ある別の学習問題に帰着することができれば，統合的に理論解析が行える．

本研究では，**学習問題の統合的理論解析を目指した帰着手法の開発**する．端的に言えば，学習問題の統合化に特化した帰着手法の開発である．学習問題を統合的に帰着できれば，今までの個別解析からの脱却が可能となり，「どのようなアルゴリズムを用いて，どれくらいの計算量で解くことができ，どのような理論性能が保証できるか」をまとめて解析することが可能である．

2. 研究成果

(1)概要

機械学習問題における，経験誤差最小化問題の一般化帰着スキームを開発した．具体的には，ある機械学習問題 A と学習問題 B があり，A と B の損失が一致するようなインスタンスのペア(入力，出力)変換関数，ならびに仮説変換関数が存在するとき，A が B に帰着可能であるという，経験誤差最小化帰着スキームを開発した．これにより，B の理論解析結果を用い，汎化性能の導出および学習アルゴリズムが即時に適用できる．マルチインスタンス学習(MIL)問題について，本帰着スキームの適用を考え，帰着可能な問題に適用可能な汎化性能と，学習アルゴリズムを示した．古典的な機械学習問題から，近年提案された機械学習問題まで，様々な学習問題が帰着可能であることを示した．具体的には，マルチクラス学習問題，マルチラベル学習問題，補ラベル学習問題，マルチタスク学習問題が MIL 問題に帰着可能であることを証明した．また，MIL への帰着スキームをもとに，新たな学習問題(最上位アイテム学習問題，negative feedback 付き最上位アイテム学習問題，perfectionistic loss に基づくマルチラベル学習)を提案し，いずれも MIL 問題に帰着可能であることを示した．補ラベル学習問題に，本帰着に基づく学習アルゴリズムを実装し，実用的にも有用であることを示した．最後に，本研究は人工知能理論トップ会議である UAI2022 で発表を行った[1]．また，本研究における直接的な成果ではないものの，本研究の活動の一部であった「様々な分野の研究者との交流を通して様々な学習問題に関する定式化と理論の広く深く理解する」活動により，署名照合問題おける上位特化学習問題に関する定式化に着手し，新たな署名照合学習アルゴリズムを開発した．また，実験結果において，署名照合分野において既存手法を上回る性能を示した．共著者らとともに国際会議 DAS2022 で発表を行い，Best Student Paper Award を受賞した[2]．

(2)詳細

本研究の目的は，大きく4つのテーマ分けられ，以下 A～D を明らかにすることである： A. MIL 問題へ帰着するための統合的帰着手法，B. MIL 問題に帰着可能な問題の具体的な性質，C. MIL への帰着の実用性，D. 統合的帰着手法の一般形の指針．

D. 統合的帰着手法の一般形の指針

機械学習問題における，経験誤差最小化問題(Empirical Risk Minimization)の一般化帰着スキーム(以下，ERM-reduction scheme)を開発した．具体的には，ある機械学習問題 A と学習問題 B があり，A と B の損失が一致するようなインスタンスのペア(入力，出力)変換関数，ならびに仮説変換関数が存在するとき，A が B に経験後最小化問題について帰着可能であるという，ERM-reduction scheme を開発した．これにより，B の理論解析結果を用い，汎化性能の導出および学習アルゴリズムが即時に適用できる．概要を次図に示す．まず，上部にある定義により ERM-reducible scheme を提案し，経験誤差最小化に関する A と B の関係，ならびに汎化性能(Rademacher complexity)に関する A と B の関係を導出した．

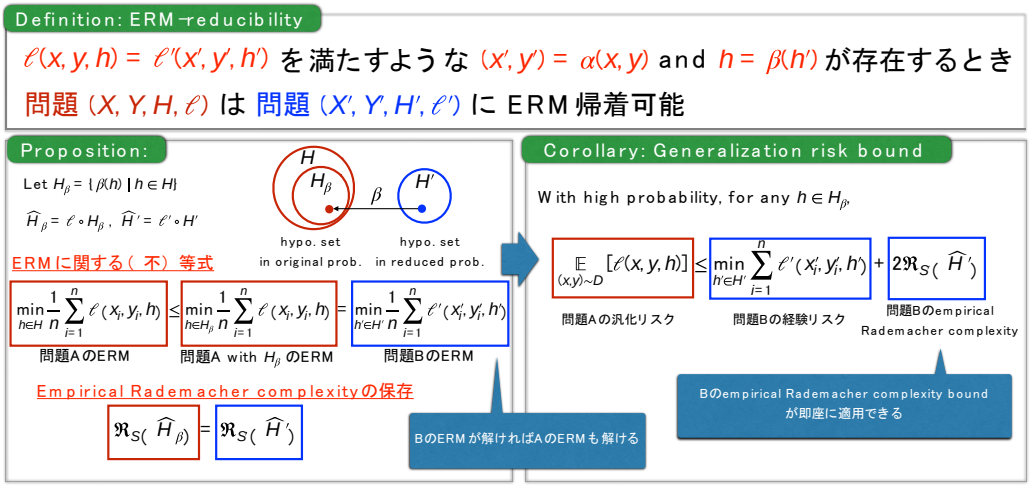
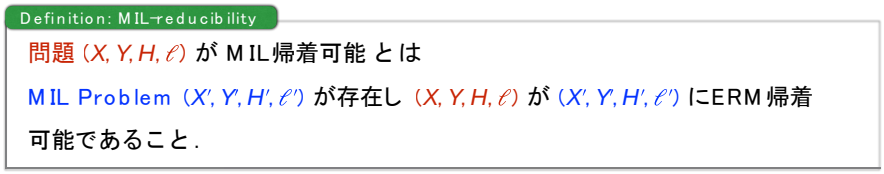


図 1: ERM-reduction scheme

A. MIL 問題へ帰着するための統合的帰着手法 および B. MIL 問題に帰着可能な問題の具体的な性質

マルチインスタンス学習（MIL）問題について，ERM-reduction scheme の適用を考え，帰着可能な問題が持つべき具体的性質，および帰着可能な問題に適用可能な汎化性能と，学習アルゴリズムを示した．概要を次図に示す．



MILの定義を踏まえると， MIL帰着可能であるための条件は:

$\ell(x,y,h) = f_1(y, \Psi_p(\{f_2(g(z)) \mid z \in x\}))$ を満たすような

$(x',y') = \alpha(x,y)$ and $h = \beta(h')$ が存在すること

図 2: MIL 帰着可能な問題の具体的性質

図 2 は，「なんらかの値の集合」を p-norm 関数で統合した式を目的関数にもつような学習問題は MIL に帰着可能であることを示唆し，実際に様々な学習問題が帰着可能である(後述)．

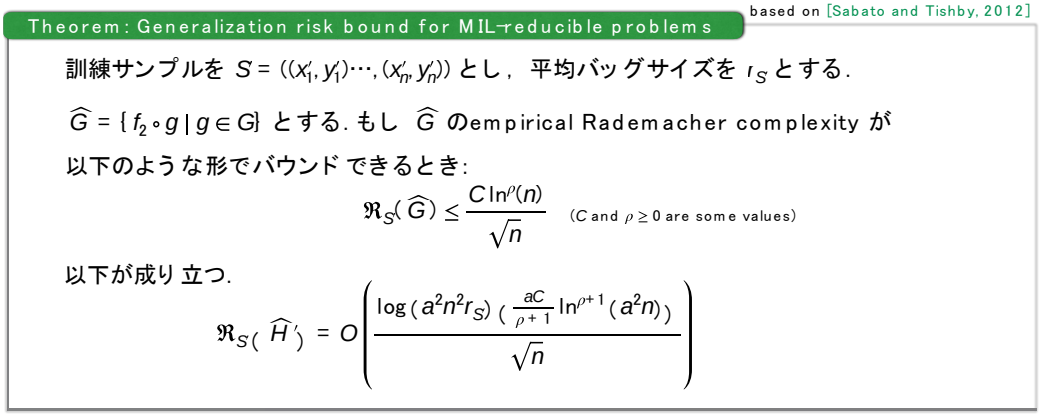


図 3: MIL 帰着可能な問題に適用できる汎化性能のバウンド

図 4: MIL 帰着可能な問題に対する最適化アルゴリズム

図 3 の結果より，従来タスクごとに解析を行って導出していた汎化誤差の導出が，帰着により容易に導出可能となる．また，図 4 の結果より，学習アルゴリズムが即座に設計でき，タスクごとにスクラッチから実装する必要がなくなる．

また，古典的な機械学習問題から，近年提案された機械学習問題まで，様々な学習問題が帰着可能であることを示した．具体的には，マルチクラス学習問題，マルチラベル学習問題，補ラベル学習問題，マルチタスク学習問題が MIL 問題に帰着可能であることを証明した．また，MIL への帰着スキームをもとに，新たな学習問題（最上位アイテム学習問題，negative feedback 付き最上位アイテム学習問題，perfectionistic loss に基づくマルチラベル学習）を提案し，いずれも MIL 問題に帰着可能であることを示した．これにより，これらの問題の汎化性能及びアルゴリズムを統合的かつ簡易的に導出可能となる．結果の概要を次図に示す．

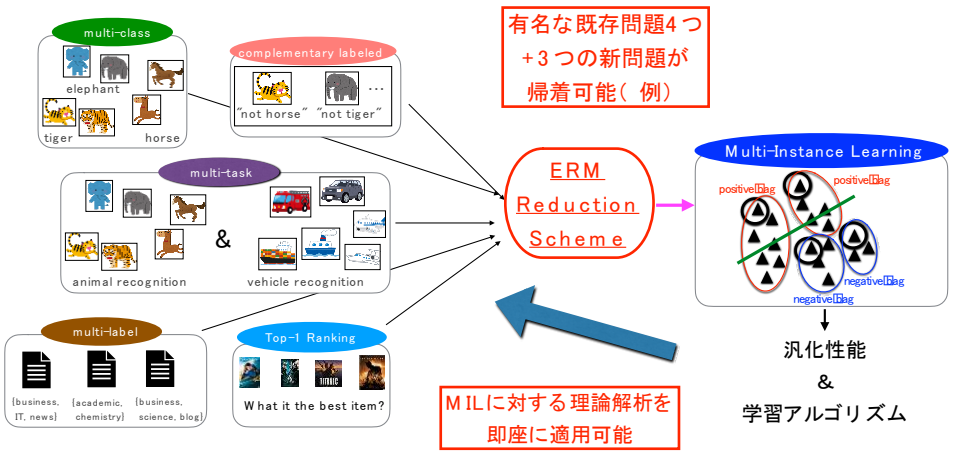


図 5: MIL 帰着可能な問題の例

C. MIL への帰着の実用性

最後に，提案した一般化帰着スキームならびに MIL への帰着の実用的な有用性を示した．具体的には，補ラベル学習問題に対し，本帰着スキームで導出した学習シナリオを適用した．人エデータおよび実データ

Dataset	Class	Dim.	Ours	Ishida+
artificial1	5	50	**0.9999**	0.9998
artificial2	10	50	**0.808**	0.646
artificial3	25	50	0.063	**0.065**
covertype	7	54	**0.562**	0.549
satimage	7	36	**0.804**	0.751
waveform	3	40	**0.833**	0.832
yeast	10	8	0.348	**0.407**

に対し，既存のアルゴリズム（Ishida et al., 2017）と比較した結果を図 3 に示す．多くのデータセットに対して高い精度を残しており，提案した帰着スキームが実用的にも有用であることが確認できた．また，導出した学習アルゴリズムが，凸計画法または DC 計画法となることから，経験誤差最小化に必要なハイパーパラメータが既存手法と比べて大きく低減できていることも示した．

3. 今後の展開

学習問題の統合的帰着による**学習問題の体系化および実用性の立証**を行っていく. 具体的には, ①ERM-reduction scheme によって導かれる学習問題のクラスを明らかにすること, および, ② ERM-reduction scheme に基づいた, 理論性能保証を見据えた学習問題の創出である. ①により, 計算量理論分野のように, 学習問題を「難しさ」によって整理することが可能となり, 従来とは異なるタスク非依存の体系化が可能となる. また, ②により, 従来の学習問題創出プロセス「タスク→定式化→理論解析」ではなく, 「タスク→理論解析付き定式化」と, 定式化の新たな道筋を示すことで, 統合的帰着スキームの実用性を示すことができる.

これらの研究成果を踏まえ, MIL への帰着を題材とし, MIL へ帰着可能な問題のクラスの特徴づけを明らかにし, 帰着に対する一般手順を明らかにするとともに, 学習問題の体系化を行う. また, 研究成果として得られた新たな学習問題を手がかりとしながら, 理論結果を見据えた新たな学習問題の創出と実アプリケーションを示す.

MIL を題材とした研究は向こう1年間程度で行い, 一般体系化の指針を示す. その後は MIL 以外の事例も試行錯誤しながら, 1～5 年程度をかけて一般体系化とその実用性を行っていく.

4. 自己評価

当初計画にあげていた目的は達成することができた. 当初スケジュールで予定していた「MIL への帰着スキーム→一般化帰着スキーム」という流れにはならず, ほぼ同時に達成することとなった. しかし, これは当初考えていたとおり, 各テーマを行き来しながら考えていくことで目的に到達できるであろうという目論見通りとなった.

また, ACT-X 関係者, 共同研究者等, 様々な研究者との交流を積極的に行い, 様々な学習問題に関する知見を得ることで, 研究を大きくすすめることができた. 具体的には, 医療情報学分野の研究者や, 文字情報学研究者らと共同研究を行い, 細胞検出問題や, 署名照合問題における上位特化学習問題に関する定式化に着手し, 新たなアルゴリズムを開発した. これらは本研究の直接的な成果ではないものの, 上位特化学習問題に関する知見は本研究における MIL への帰着事例創出に大きく貢献した.

本研究は, あらゆる学習問題に跨っている研究であることから, 領域会議や国内外の学会を通し, ACT-X 内外の研究者らと議論を交わすことができた. 今後も, 本研究を軸として様々な研究者との交流を積極的に行い, 学習問題の定式化および理論構築の知見で異分野に貢献しながら, 得られた知見を自身の研究に還元して新たな成果を求めていく.

5.　主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：5件

1. **Daiki Suehiro**, Eiji Takimoto, "Simplified and Unified Analysis of Various Learning Problems by Reduction to Multiple-Instance Learning", Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022), 2022.8, PMLR 180:1896-1906.
経験誤差最小化問題に基づく学習問題の帰着スキームを提案した．帰着スキームを用いることで，汎化誤差の導出およびアルゴリズムの設計が容易となる．また，帰着スキームをマルチインスタンス学習問題に適用し，様々な学習問題がマルチインスタンス学習問題に帰着できることを示した．
2. Xiaotong Ji, Yan Zheng, **Daiki Suehiro** and Seiichi Uchida, "Revealing Reliable Signatures by Learning Top-Rank Pairs", Proceedings of the 15th IAPR International Workshop on Document Analysis Systems, pp.323-337, 2022.5.
署名照合問題に対し，上位特化学習を適用するための定式化を考案した．単純な精度だけでなく，従来本問題で軽視されていた予測の「信頼性」すなわち「本物の署名であることに対する信頼度」を大きくするための学習を可能とした．従来手法よりも高い性能で署名照合が行えることを示した．
3. Kazuma Fujii, **Daiki Suehiro**, Kazuya Nishimura, and Ryoma Bise, "Cell Detection from Imperfect Annotation by Pseudo Label Selection Using P-classification", Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI2021), pp. 425--434.
細胞検出タスクにおいて，少ないアノテーションデータから検出器を学習するためのアルゴリズムを開発した．部分教師データからの学習手法と上位特化学習を組み合わせ，新ライン度の高い「疑似教師データ」を与える手法であり，従来手法よりも高い精度での検知が可能となった．

（2）特許出願

研究期間全出願件数：0 件（特許公開前のものも含む）

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

1. DAS2022 Best Student Paper Awards
2. 第 46 回情報論的学習理論と機械学習研究会（IBISML） 口頭発表（査読なし）